



# UNITED STATES PATENT AND TRADEMARK OFFICE

*nn*  
UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/811,962	03/30/2004	Akihiro Okumura	32178-201736	2884
26694	7590	07/13/2007		
VENABLE LLP P.O. BOX 34385 WASHINGTON, DC 20043-9998			EXAMINER PHAM, MICHAEL	
			ART UNIT	PAPER NUMBER
			2167	
			MAIL DATE	DELIVERY MODE
			07/13/2007	PAPER

**Please find below and/or attached an Office communication concerning this application or proceeding.**

The time period for reply, if any, is set in the attached communication.

<b>Office Action Summary</b>	Application No. 10/811,962	Applicant(s) OKUMURA ET. AL.	
	Examiner Michael D. Pham	Art Unit 2167	

**-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --**

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

- 1) ☒ Responsive to communication(s) filed on 06 April 2007.
- 2a) ☒ This action is **FINAL**.                      2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

- 4) ☒ Claim(s) 1-8, 10-46 and 48-65 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-8, 10-46 and 48-65 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

**Application Papers**

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on \_\_\_\_\_ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All    b) ☐ Some \*    c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

- |  |   |
|--|---|
| 1) <input type="checkbox"/> Notice of References Cited (PTO-892)   | 4) <input type="checkbox"/> Interview Summary (PTO-413)<br>Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)                       | 5) <input type="checkbox"/> Notice of Informal Patent Application                       |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08)<br>Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____  |

## DETAILED ACTION

### *Claim status*

1. Claims 1-6, 10-15, 17-25, 28-36, 39-44, 48-53, 56-60, and 63-65 are currently amended.
2. Claims 9 and 47 are canceled.
3. Claims 1-8, 10-46, and 48-65 are pending.

### *Claim Objections*

4. Claim 2-6, 10-15, 17-25, 28-36, 40-44, 48-53, 56-60, and 63-65 are objected to because of the following informalities: Claims when amended contain non-underlined portions where an underline was needed. Example, prior claims such as claim 2 contained "a extracting unit"; while the newly amended claims underlined "wherein" however did not underline "the".

However this just appears to be a typo. In order to expedite the case, it will be treated as such for this office action. Please acknowledge if this is accurate.

### *Claim Rejections - 35 USC § 112*

5. Prior rejections under 35 U.S.C. 112 are respectfully withdrawn for claims 5, 13, 24, 35, 43, 51, and 59.

### *Claim Rejections - 35 USC § 103*

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person

Art Unit: 2167

having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

**6. Claims 1, 2, 7 – 10, and 15-21 are rejected under 35 U.S.C. 103(a) as being unpatentable over Pirolli et al. (hereinafter Pirolli, US 5,895,470) further in view of Tsuda (hereafter Tsuda, US 7,003,442).**

7. Regarding claim 1, **Pirolli** discloses an information extracting apparatus for extracting designated information from a document group having a hypertext structure in which documents are mutually related by link information (See column 6, lines 8 – 10 “Referring to FIG. 2, the walker uses the Hypertext Transfer Protocol (HTTP) to request and retrieve a web page, step 201.”), comprising:

a start point address designating unit [walker] which designates an address of the document serving as a start point where said information is extracted (See column 6, lines 4-7 “The site’s topology is ascertained via ‘the walker’, an autonomous agent that, given a starting point, performs an exhaustive breadth-first traversal of pages within the web locality.” The start point addressing unit is defined in the specification in paragraph [0068] as allowing the user to designate the address of a target document to be extracted, which is what is occurring here.); and

a category designating unit which designates a category of the information to be extracted (See column 8, lines 55 – 58 “These functional categories might be defined by a user’s specific set of interests, or the categories might be extracted from the collection itself through inductive techniques.”);

an extracting unit which:

Art Unit: 2167

extracts the information corresponding to said category from the target document designated by said start point address designating unit (See column 6, lines 15 – 19 “The meta-information for the page is also extracted and stored, step 204. The meta-information includes at least the following page meta-information: name, title, list of children (pages associated by hyperlinks), file size, and the time the page was last modified.”) and, if the information corresponding to said category could not be extracted from said target document, extracts said information from the related document of said target document on the basis of the address of said document. (See column 6, lines 24 – 26 “The list of pages to request and retrieve is then used to obtain the next page, step 206. The process then repeats per step 202 until all of the pages on the list have been retrieved.”)

**Pirolli** does not explicitly disclose

a category layer specifying unit in which the category of the information to be extracted is expressed by a layer structure;

an extracting unit which:

in case where only an extraction result of a lower layer in said layer structure exists and an extraction result of an upper layer is missing as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, extracts a character string of a layer which is higher than that of the extraction result lower layer from the related document of said target document.

On the other hand, **Tsuda** discloses

Art Unit: 2167

a category layer specifying unit in which the category of the information to be extracted is expressed by a layer structure (col. 7 lines 15-17, discloses “a directory file creating unit creates hyper text format directory file with the data 56, the data 33, the data 52, the data 53, the data 54, and the data 55”);

an extracting unit which:

in the case where only an extraction result of a lower layer in said layer structure exists and an extraction result of an upper layer is missing as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, extracts a character string of a layer which is higher than that of the extraction result lower layer from the related document of said target document (col. 15 lines 21-23 that “the directory file creating unit designates a set of keywords registered in the field down of the keyword w of the keyword table 62 as S2.” Col. 15 lines 24-29, “the directory file creating unit 43 determines whether or not s2 is empty. When s2 is not empty the directory file creating unit extracts a keyword from s2. Next the directory file creating unit determines whether or not the path field of the keyword u is empty.”)

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli** with that of **Tsuda** because both are related to organized linked documents and by including the extraction method as disclosed in **Tsuda**, the apparatus can effectively search multiple pages and combine the results obtained over multiple pages of the same document. It is for this reason that one of ordinary skill in the art would have been motivated to include a category layer specifying unit in which the category of the information to

Art Unit: 2167

be extracted is expressed by a layer structure; an extracting unit which, in the case where only an extraction result of a lower layer in said layer structure exists and an extraction result of an upper layer is missing as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, extracts a character string of a layer which is higher than that of the extraction result lower layer from the related document of said target document.

8. Regarding claim 2, **Pirolli** additionally discloses wherein extracting unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

9. Regarding claim 7, **Pirolli** additionally teaches said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. (See column 6, lines 24-26 “The list of pages to request and retrieve is then used to obtain the next page, step 206.” These are examples of link destination documents included in the related document.)

10. Regarding claim 8, **Pirolli** additionally teaches said upper document [returned page] is at least either a document of a specific name existing in a one-upper directory of the target document or a link source document existing in the one-upper directory. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the returned page is a source of links in an upper directory to the pages in which the links are directed.)

11. Regarding claim 10, **Pirolli** additionally discloses wherein extracting unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

12. Regarding claim 15, **Pirolli** additionally teaches said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. (See column 6, lines 24-26 “The list of pages to request and retrieve is then used to obtain the next page, step 206.” These are examples of link destination documents included in the related document.)



Art Unit: 2167

13. Regarding claim 16, **Pirolli** additionally teaches said upper document [returned page] is at least either a document of a specific name existing in a one-upper directory of the target document or a link source document existing in the one-upper directory. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the returned page is a source of links in an upper directory to the pages in which the links are directed.)

14. Regarding claim 17, **Pirolli** teaches a method substantially as claimed. **Pirolli** fails to explicitly teach a processing unit which outputs a character string, as an extraction result, obtained by synthesizing the extraction result of said lower layer and the extraction result of said upper layer.

However, **Tsuda** teaches

a processing unit which outputs a character string, as an extraction result, obtained by synthesizing the extraction result of said lower layer and the extraction result of said upper layer. (See column 19, lines 34 – 35 “The outputting unit 164 is used to display query messages to the user and processed results.”)

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli** with that of **Tsuda** because both are related to organized linked documents and by including the extraction method as disclosed in **Tsuda**, the apparatus can effectively search multiple pages and combine the results obtained over multiple pages of the same document. It is for this reason that one of ordinary skill in the art would have

Art Unit: 2167

been motivated to include a category layer specifying unit in which the category of the information to be extracted is expressed by a layer structure; an extracting unit which, in the case where only an extraction result of a lower layer in said layer structure exists and an extraction result of an upper layer is missing as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, extracts a character string of a layer which is higher than that of the extraction result of said lower layer from the related document of said target document; and a processing unit which outputs a character string, as an extraction result, obtained by synthesizing the extraction result of said lower layer and the extraction result of said upper layer.

15. Regarding claim 18, **Pirolli** teaches a method substantially as claimed. **Pirolli** fails to explicitly teach wherein the processing unit has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule. However **Tsuda** teaches wherein the processing unit has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule. (See column 10, lines 6 – 10 “The merger 84 merges the hierarchical relation 32, the character sub-string relation 85, and the hierarchical relation generated by the rule evaluating unit 83 and generates the hierarchical relation.”) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli** with that of **Tsuda** because both are related to organized linked documents and by including the synthesizing rule as disclosed in

Art Unit: 2167

**Tsuda**, the apparatus can effectively combine the results obtained over multiple pages of the same document. It is for this reason that one of ordinary skill in the art would have been motivated to include teach wherein the processing unit has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule.

16. Regarding claim 19, **Pirolli** teaches a method substantially as claimed. **Pirolli** fails to explicitly teach wherein the processing unit forms the character string of the processing result by coupling a plurality of character strings in order from the extraction result of the upper layer to the extraction result of the lower layer on the basis of the layer structure. However, **Tsuda** teaches wherein the processing unit forms the character string [keyword] of the processing result by coupling a plurality of character strings in order from the extraction result of the upper layer to the extraction result of the lower layer on the basis of the layer structure. (See column 18, lines 5 – 8 “The processing unit 121 comprises a keyword trimming unit, a keyword relation extracting unit, a directory file creating unit, a searching unit, and a www sever.” And see column 7, lines 50 – 55 “A keyword able contains combinations of [keyword ID (KID), keyword, reading information a set of higher word Ids (UP); a set of lower word Ids (DOWN), a set of associative word Ids (Rel), a set of equivalent keyword Ids (Ea), a path, a new word flag (new)].”) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli** with that of **Tsuda** because both are related to organized linked documents and by including the coupling of the strings as disclosed in **Tsuda**, the apparatus can effectively combine the results obtained over multiple pages of the same

document. It is for this reason that one of ordinary skill in the art would have been motivated to include wherein the processing unit forms the character string of the processing result by coupling a plurality of character strings in order from the extraction result of the upper layer to the extraction result of the lower layer on the basis of the layer structure.

17. Regarding claim 20, **Pirolli** teaches a method substantially as claimed. **Pirolli** fails to explicitly teach wherein the processing unit has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule. However **Tsuda** teaches wherein the processing unit has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule. (See column 10, lines 6 – 10 “The merger 84 merges the hierarchical relation 32, the character sub-string relation 85, and the hierarchical relation generated by the rule evaluating unit 83 and generates the hierarchical relation.”) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli** with that of **Tsuda** because both are related to organized linked documents and by including the synthesizing rule as disclosed in **Tsuda**, the apparatus can effectively combine the results obtained over multiple pages of the same document. It is for this reason that one of ordinary skill in the art would have been motivated to include teach wherein the processing unit has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule.

18. Regarding claim 21, **Pirolli** additionally discloses wherein extracting unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

19. **Claim 39, 55, and 63-65 are rejected under 35 U.S.C. 103(a) as being unpatentable over U.S. Patent Application Publication 2004/0019499 by Murashita (hereafter Murashita) further in view of Tsuda (hereafter Tsuda, US 7,003,442).**

20. Regarding claim 39, **Murashita** discloses an information extracting apparatus for extracting designated information from a document group having a hypertext structure in which documents are mutually related by link information (See page 1, paragraph [0008] “The search engine is a system for registering the document on the Internet and its keyword into a server and a searching information by a keyword inputted by the user and is called an agent, an automatic collecting robot, or the like. The search engine scans the document stored in the server on the Internet and forms a document for displaying and a keyword database for searching.”), comprising:

an extracting unit [information collection apparatus] which extracts target information from said document group and, in the case where addition or updating of a document occurs for said document group, executes an extracting process to which such addition or updating is reflected each time said addition or updating occurs, and outputs an extraction result including said target information and its document address (See page 9, paragraph [0167] “As mentioned above, in the information collecting apparatus of the invention, the specific site is monitored as an event collecting destination site, if the information in this event collecting destination site has been updated, the keyword to specify the event such as announcement of a new product, incidence of the new virus, or the like is formed from contents of the update, and the information including the keyword is collected from the information collecting destination site by the keyword.”);

an extraction result storing unit which stores the extraction result from said extracting unit as extraction result information (see page 9, paragraph [0171] “In step s11, the documents obtained by the information searching unit 26 by using the keyword are stored in the document storing unit.”);

a start point address designating unit which designates an address of a document serving as a start point where said designated information is extracted (See page 9, paragraph [0165] “If the user wants to collect information regarding a computer virus by using the information, in step S1, a URL of an antivirus software developing company is preliminarily registered into the event collecting destination site.”); and

a searching unit which:

extracts information from the target document of the document address designated by said start point address designating unit and its related document with reference to the extraction result information in said extraction result storing unit (See page 9, paragraph [0166] "...the useful information showing how to cope with the new virus as a user of the personal computer is automatically collected by the search of the information collecting destination site by the keyword such as a virus name or the like extracted by the detection of the incidence of the new virus, and it can be shown to the user.").

a category [keyword] designating unit which designates a category of the information to be extracted (See page 9, paragraph [0167] "... if the information in this event collecting destination site has been updated, the keyword to specify the event such as announcement of a new product, incidence of the new virus, or the like is formed from contents of the update, and the information including the keyword is collected from the information collecting destination site by the keyword.")

a search unit which:

extracts the information belonging to the category designated by said designating unit (See page 9, paragraph [0166] "...the useful information showing how to cope with the new virus as a user of the personal computer is automatically collected by the search of the information collecting destination site by the keyword such as a virus name or the like extracted by the detection of the incidence of the new virus, and it can be shown to the user.").

However, **Murashita** does not explicitly disclose

Art Unit: 2167

a category layer specifying unit in which the category of the information to be extracted is expressed by a layer structure; and

a search unit which:

in the case where an extraction result of an upper layer is missing only in an extraction result of a lower layer in said layer structure as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, extracts a character string of a layer which is higher than that of the extraction result of said lower layer from the related document of said target document.

On the other hand, **Tsuda** discloses

a category layer specifying unit in which the category of the information to be extracted is expressed by a layer structure (col. 7 lines 15-17, discloses “a directory file creating unit creates hyper text format directory file with the data 56, the data 33, the data 52, the data 53, the data 54, and the data 55”);

a search unit which:

in the case where an extraction result of an upper layer is missing only in an extraction result of a lower layer in said layer structure as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, extracts a character string of a layer which is higher than that of the extraction result of said lower layer from the related document of said target document (col. 15 lines 21-23 that “the directory file creating unit designates a set of keywords registered in the field down of the keyword w of the keyword table 62 as S2.” Col. 15 lines 24-29, “the directory



Art Unit: 2167

file creating unit 43 determines whether or not s2 is empty. When s2 is not empty the directory file creating unit extracts a keyword from s2. Next the directory file creating unit determines whether or not the path field of the keyword u is empty.”)

21. Regarding claim 55, **Murashita** teaches an apparatus substantially as claimed.

**Murashita** does not explicitly the searching unit outputs a character string, as an extraction result, obtained by synthesizing the extraction result of said lower layer and the extraction result of said upper layer.

However, **Tsuda** discloses the searching unit outputs a character string, as an extraction result, obtained by synthesizing the extraction result of said lower layer and the extraction result of said upper layer. (See column 15, lines 24 – 29 “Next, the directory file creating unit 43 determines whether or not s2 is empty at step s76). When s2 is not empty the directory file creating unit extracts a keyword from s2. Next the directory file creating unit determines whether or not the path field of the keyword u is empty.”)

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita** with that of **Tsuda** because they are both related to hypertext document organization and by including the concept of extracting from different layers as disclosed in **Tsuda**, the apparatus the apparatus can effectively search multiple pages and combine the results obtained over multiple pages of the same document. It is for this reason that one of ordinary skill in the art would have been motivated to include the searching unit outputs a character string, as an extraction result, obtained by synthesizing the extraction result of said lower layer and the extraction result of said upper layer.

22. Regarding claim 63, the combination of **Murashita** and **Tsuda** teaches wherein the searching unit has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule. (See **Tsuda** column 10, lines 6 – 10 “The merger 84 merges the hierarchical relation 32, the character sub-string relation 85, and the hierarchical relation generated by the rule evaluating unit 83 and generates the hierarchical relation.”)

23. Regarding claim 64, the combination of **Murashita** and **Tsuda** teaches wherein the searching unit forms a character string [keyword] of a processing result by coupling a plurality of character strings in order from the extraction result of the upper layer to the extraction result of the lower layer on the basis of the layer structure. (See **Tsuda** column 18, lines 5 – 8 “The processing unit 121 comprises a keyword trimming unit, a keyword relation extracting unit, a directory file creating unit, a searching unit, and a www sever.” And see column 7, lines 50 – 55 “A keyword able contains combinations of [keyword ID (KID), keyword, reading information a set of higher word Ids (UP), a set of lower word Ids (DOWN), a set of associative word Ids (Rel), a set of equivalent keyword Ids (Ea), a path, a new word flag (new)].”)

24. Regarding claim 65, the combination of **Murashita** and **Tsuda** teaches wherein the searching unit has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing

Art Unit: 2167

result in accordance with said synthesizing rule. (See **Tsuda** column 10, lines 6 – 10 “The merger 84 merges the hierarchical relation 32, the character sub-string relation 85, and the hierarchical relation generated by the rule evaluating unit 83 and generates the hierarchical relation.”)

25. **Claims 3–6, 11-14, and 22-27 are rejected under 35 U.S.C. 103(a) as being unpatentable over Pirolli and Tsuda as applied to claim 1 above, and further in view of Sweet et al. (hereinafter Sweet, US 2002/0073074).**

26. Regarding claim 3, **Pirolli and Tsuda** teach an information extracting apparatus substantially as claimed. **Pirolli and Tsuda** do not explicitly teach a maximum link depth designating unit which designates a maximum link depth; and wherein extracting unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

However, **Sweet** teaches a maximum link depth designating unit which designates a maximum link depth; and wherein extracting unit which, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth. (See page 6, paragraph [0063] “One web traversal criterion which may be specified

Art Unit: 2167

by the user is a maximum depth criterion. This criterion limits the depth of recursive calls to FetchAndIncorporate, and thus limits the 'link distance' between the initially retrieved document and subsequently retrieved documents to be incorporated into the target document.”)

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli and Tsuda** with that of **Sweet** because all are related to operating on linked documents and by including the maximum link depth as disclosed in **Sweet**, the apparatus can remain efficient by having a limit on the recursion, rather than having unlimited recursion. It is for this reason that one of ordinary skill in the art would have been motivated to include a maximum link depth designating unit which designates a maximum link depth; and wherein extracting unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

27. Regarding claim 4, **Pirolli** additionally discloses wherein extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

28. Regarding claim 5, **Pirolli** additionally discloses wherein extracting unit which executes the information extracting process in order of the document in which a value of the link depth is small. (See column 6, lines 12 – 26 where the hypertext links are extracting at the higher document depth first, then the links on those pages are executed, finding larger depth value links and then repeating. In other words, the executing starts with a smaller link depth and then goes to larger link depths during the extraction process.)

29. Regarding claim 6, **Pirolli** additionally discloses wherein extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

30. Regarding claim 11, **Pirolli** and **Tsuda** teach an information extracting apparatus substantially as claimed. **Pirolli** and **Tsuda** does not explicitly teach a maximum link depth designating unit which designates a maximum link depth; and wherein extracting unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

However, **Sweet** teaches a maximum link depth designating unit which designates a maximum link depth; and wherein extracting unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth. (See page 6, paragraph [0063] “One web traversal criterion which may be specified by the user is a maximum depth criterion. This criterion limits the depth of recursive calls to FetchAndIncorporate, and thus limits the ‘link distance’ between the initially retrieved document and subsequently retrieved documents to be incorporated into the target document.”)

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli** and **Tsuda** with that of **Sweet** because all are related to operating on linked documents and by including the maximum link depth as disclosed in **Sweet**, the apparatus can remain efficient by having a limit on the recursion, rather than having unlimited recursion. It is for this reason that one of ordinary skill in the art would have been motivated to include a maximum link depth designating unit which designates a maximum link depth; and wherein extracting unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

31. Regarding claim 12, **Pirolli** additionally discloses wherein extracting unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other

pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

32. Regarding claim 13, **Pirolli** additionally discloses wherein extracting unit executes the information extracting process in order of the document in which a value of the link depth is small. (See column 6, lines 12 – 26 where the hypertext links are extracting at the higher document depth first, then the links on those pages are executed, finding larger depth value links and then repeating. In other words, the executing starts with a smaller link depth and then goes to larger link depths during the extraction process.)

33. Regarding claim 14, **Pirolli** additionally discloses wherein extracting unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

34. Regarding claim 22, **Pirolli** and **Tsuda** teach an information extracting apparatus substantially as claimed. **Pirolli** and **Tsuda** do not explicitly teach a maximum link depth designating unit which designates a maximum link depth; and wherein the extracting unit, in the

case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

However, **Sweet** teaches a maximum link depth designating unit which designates a maximum link depth; and wherein the extracting unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth. (See page 6, paragraph [0063] "One web traversal criterion which may be specified by the user is a maximum depth criterion. This criterion limits the depth of recursive calls to FetchAndIncorporate, and thus limits the 'link distance' between the initially retrieved document and subsequently retrieved documents to be incorporated into the target document.")

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli** and **Tsuda** with that of **Sweet** because the references are related to operating on linked documents and by including the maximum link depth as disclosed in **Sweet**, the apparatus can remain efficient by having a limit on the recursion, rather than having unlimited recursion. It is for this reason that one of ordinary skill in the art would have been motivated to include a maximum link depth designating unit which designates a maximum link depth; and wherein the extracting unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.



35. Regarding claim 23, **Pirolli** additionally discloses wherein the extracting unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

36. Regarding claim 24, **Pirolli** additionally discloses wherein the extracting unit executes the information extracting process in order of the document in which a value of the link depth is 3 or fewer. (See column 6, lines 6 breadth first, e.g. link depth starts at root and continues by level.)

37. Regarding claim 25, **Pirolli** additionally discloses wherein the extracting unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

38. Regarding claim 26, **Pirolli** additionally teaches said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. (See column 6, lines 24-26 “The list of pages to request and retrieve is then used to obtain the next page, step 206.” These are examples of link destination documents included in the related document.)

39. Regarding claim 27, **Pirolli** additionally teaches said upper document [returned page] is at least either a document of a specific name existing in a one-upper directory of the target document or a link source document existing in the one-upper directory. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the returned page is a source of links in an upper directory to the pages in which the links are directed.)

40. **Claims 28 – 32 and 37 – 38 are rejected under 35 U.S.C. 103(a) as being unpatentable over Pirolli in view of Tsuda as applied to claim 17 above, and further in view of Kunitake et al. (hereinafter Kunitake, US 2002/0073074).**

41. Regarding claim 28, **Pirolli** and **Tsuda** teach an information extracting apparatus substantially as claimed. **Pirolli** and **Tsuda** do not explicitly teach wherein the extracting unit, in the case where the extraction result is separated into a plurality of character strings of the extraction result of the lower layer and the extraction result of the upper layer in said layer

structure as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, outputs said plurality of character strings as an extraction result of the lower layer and an extraction result of the upper layer.

However, **Kunitake** teaches wherein the extracting unit, in the case where the extraction result is separated into a plurality of character strings [instruction strings] of the extraction result of the lower layer and the extraction result of the upper layer in said layer structure as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, outputs said plurality of character strings as an extraction result [document processing description] of the lower layer and an extraction result of the upper layer. (See page 12, paragraph [0306] "Next, a document processing description synthesizing unit inputs instruction strings separated from plural original documents or templates, merges and sorts the instruction strings, and outputs a document processing description after conversion and synthesis.")

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli** and **Tsuda** with that of **Kunitake** because the references are related to operating on linked documents and by including the character strings as disclosed in **Kunitake**, the apparatus can combine information from various layers of the document all in one result. It is for this reason that one of ordinary skill in the art would have been motivated to include wherein the extracting unit, in the case where the extraction result is separated into a plurality of character strings of the extraction result of the lower layer and the extraction result of the upper layer in said layer structure as a result of the extraction of the

Art Unit: 2167

information corresponding to the category from the target document designated by said start point address designating unit, outputs said plurality of character strings as an extraction result of the lower layer and an extraction result of the upper layer.

42. Regarding claim 29, the combination of **Pirolli**, **Tsuda**, and **Kunitake** teaches wherein the processing unit has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule. (See **Tsuda** column 10, lines 6 – 10 “The merger 84 merges the hierarchical relation 32, the character sub-string relation 85, and the hierarchical relation generated by the rule evaluating unit 83 and generates the hierarchical relation.”)

43. Regarding claim 30, The combination of **Pirolli**, **Tsuda**, and **Kunitake** teaches wherein the processing unit forms the character string [keyword] of the processing result by coupling a plurality of character strings in order from the extraction result of the upper layer to the extraction result of the lower layer on the basis of the layer structure. (See **Tsuda** column 18, lines 5 – 8 “The processing unit 121 comprises a keyword trimming unit, a keyword relation extracting unit, a directory file creating unit, a searching unit, and a www sever.” And see column 7, lines 50 – 55 “A keyword able contains combinations of [keyword ID (KID), keyword, reading information a set of higher word Ids (UP), a set of lower word Ids (DOWN), a set of associative word Ids (Rel), a set of equivalent keyword Ids (Ea), a path, a new word flag (new)].”)

44. Regarding claim 31, the combination of **Pirolli**, **Tsuda**, and **Kunitake** teaches wherein the processing unit which has a predetermined synthesizing rule in the case of synthesizing a plurality of character strings expressed by the layer structure and forms a character string of a processing result in accordance with said synthesizing rule. (See **Tsuda** column 10, lines 6 – 10 “The merger 84 merges the hierarchical relation 32, the character sub-string relation 85, and the hierarchical relation generated by the rule evaluating unit 83 and generates the hierarchical relation.”)

45. Regarding claim 32, the combination of **Pirolli**, **Tsuda**, and **Kunitake** additionally discloses wherein the extracting unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

46. Regarding claim 37, the combination of **Pirolli**, **Tsuda**, and **Kunitake** additionally teaches said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. (See **Pirolli**, column 6, lines 24-26

Art Unit: 2167

“The list of pages to request and retrieve is then used to obtain the next page, step 206.” These are examples of link destination documents included in the related document.)

47. Regarding claim 38, the combination of **Pirolli, Tsuda, and Kunitake** additionally teaches said upper document [returned page] is at least either a document of a specific name existing in a one-upper directory of the target document or a link source document existing in the one-upper directory. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the returned page is a source of links in an upper directory to the pages in which the links are directed.)

48. **Claims 33 – 36 are rejected under 35 U.S.C. 103(a) as being unpatentable over Pirolli, Tsuda, Kunitake as applied to claim 28 above, and further in view of Sweet.**

49. Regarding claim 33, **Pirolli, Tsuda, and Kunitake** teach an information extracting apparatus substantially as claimed. **Pirolli, Tsuda, and Kunitake** do not explicitly teach a maximum link depth designating unit which designates a maximum link depth; and wherein the extracting unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

However, **Sweet** teaches a maximum link depth designating unit which designates a maximum link depth; and wherein the extracting unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the

Art Unit: 2167

information from the related document of said document in a range of said designated maximum link depth. (See page 6, paragraph [0063] “One web traversal criterion which may be specified by the user is a maximum depth criterion. This criterion limits the depth of recursive calls to FetchAndIncorporate, and thus limits the ‘link distance’ between the initially retrieved document and subsequently retrieved documents to be incorporated into the target document.”)

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Pirolli**, **Tsuda**, and **Kunitake** with that of **Sweet** because the references are related to operating on linked documents and by including the maximum link depth as disclosed in **Sweet**, the apparatus can remain efficient by having a limit on the recursion, rather than having unlimited recursion. It is for this reason that one of ordinary skill in the art would have been motivated to include a maximum link depth designating unit which designates a maximum link depth; and wherein the extracting unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

50. Regarding claim 34, the combination of **Pirolli**, **Tsuda**, **Kunitake**, and **Sweet** additionally discloses an extracting unit which discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the

Art Unit: 2167

pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

51. Regarding claim 35, the combination of **Pirolli, Tsuda, Kunitake, and Sweet** additionally discloses wherein the extracting unit executes the information extracting process in order of the document in which a value of the link depth is small. (See **Pirolli**, column 6, lines 12 – 26 where the hypertext links are extracting at the higher document depth first, then the links on those pages are executed, finding larger depth value links and then repeating. In other words, the executing starts with a smaller link depth and then goes to larger link depths during the extraction process.)

52. Regarding claim 36, the combination of **Pirolli, Tsuda, Kunitake, and Sweet** additionally discloses wherein the extracting unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)



53. Claims 40, 45-46, 53-54, 56, and 61-62 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Murashita and Tsuda** further in view of **Pirolli et al. (hereinafter Pirolli, US 5,895,470)**.

54. Regarding claim 40, **Murashita and Tsuda** teaches an apparatus substantially as claimed. **Murashita and Tsuda** do not explicitly disclose wherein the searching unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. However, **Pirolli** teaches wherein the searching unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita and Tsuda** with that of **Pirolli** because all are related to information collecting from hypertext documents and by including the internal and external link discrimination as disclosed in **Pirolli**, the apparatus can be more efficient by only including the pages to search that are likely to be relevant. It is for this reason that one of ordinary skill in the art would have been motivated to include wherein the searching unit discriminates an internal link and an external link on the basis

of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

55. Regarding claim 45, **Murashita and Tsuda** teaches an apparatus substantially as claimed. **Murashita and Tsuda** do not explicitly disclose said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. However, **Pirolli** teaches said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. (See column 6, lines 24-26 “The list of pages to request and retrieve is then used to obtain the next page, step 206.” These are examples of link destination documents included in the related document.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita and Tsuda** with that of **Pirolli** because all are related to information collecting from hypertext documents and by including the types of documents as disclosed in **Pirolli**, the apparatus can search both upper and lower level documents. It is for this reason that one of ordinary skill in the art would have been motivated to include said related document includes at least one of a link destination document, a link source document, and an upper document of the target document.

56. Regarding claim 46, the combination of **Murashita, Tsuda, and Pirolli** teaches said upper document is at least either a document of a specific name existing in a one-upper directory of the target document or a link source document existing in the one-upper directory. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other

Art Unit: 2167

pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the returned page is a source of links in an upper directory to the pages in which the links are directed.)

57. Regarding claim 48, **Murashita and Tsuda** teaches an apparatus substantially as claimed. **Murashita and Tsuda** does not explicitly disclose wherein the searching unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. However, **Pirolli** teaches wherein the searching unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita and Tsuda** with that of **Pirolli** because all are related to information collecting from hypertext documents and by including the internal and external link discrimination as disclosed in **Pirolli**, the apparatus can be more efficient by only including the pages to search that are likely to be relevant. It is for this reason that one of ordinary skill in the art would have been motivated to include wherein the searching unit discriminates an internal link and an external link on the basis

of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

58. Regarding claim 53, **Murashita and Tsuda** teach an apparatus substantially as claimed. **Murashita and Tsuda** do not explicitly disclose said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. However, **Pirolli** teaches said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. (See column 6, lines 24-26: "The list of pages to request and retrieve is then used to obtain the next page, step 206." These are examples of link destination documents included in the related document.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita and Tsuda** with that of **Pirolli** because all are related to information collecting from hypertext documents and by including the types of documents as disclosed in **Pirolli**, the apparatus can search both upper and lower level documents. It is for this reason that one of ordinary skill in the art would have been motivated to include said related document includes at least one of a link destination document, a link source document, and an upper document of the target document.

59. Regarding claim 54, the combination of **Murashita, Tsuda** and **Pirolli** teaches said upper document is at least either a document of a specific name existing in a one-upper directory of the target document or a link source document existing in the one-upper directory. (See **Pirolli**, column 6, lines 12 – 15 "The returned page is then parsed to extract hyperlinks to other

Art Unit: 2167

pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the returned page is a source of links in an upper directory to the pages in which the links are directed.)

60. Regarding claim 56, **Murashita** and **Tsuda** teach an apparatus substantially as claimed. **Murashita** and **Tsuda** do not explicitly disclose wherein the searching unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. However, **Pirolli** teaches wherein the searching unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita**, **Tsuda** and **Pirolli** because they are related to operating on web documents and by including link discriminating as disclosed in **Pirolli**, the apparatus can be more efficient by only including the pages to search that are likely to be relevant. It is for this reason that one of ordinary skill in the art would have been motivated to include wherein the searching unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

61. Regarding claim 61, **Murashita** and **Tsuda** teach an apparatus substantially as claimed. **Murashita** and **Tsuda** do not explicitly disclose said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. However, **Pirolli** teaches said related document includes at least one of a link destination document, a link source document, and an upper document of the target document. (See column 6, lines 24-26 “The list of pages to request and retrieve is then used to obtain the next page, step 206.” These are examples of link destination documents included in the related document.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita** and **Tsuda** with that of **Pirolli** because they are related to information collecting from hypertext documents and by including the types of documents as disclosed in **Pirolli**, the apparatus can search both upper and lower level documents. It is for this reason that one of ordinary skill in the art would have been motivated to include said related document includes at least one of a link destination document, a link source document, and an upper document of the target document.

62. Regarding claim 62, the combination of **Murashita**, **Tsuda** and **Pirolli** teaches said upper document is at least either a document of a specific name existing in a one-upper directory of the target document or a link source document existing in the one-upper directory. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to

request and retrieve.” Here, the returned page is a source of links in an upper directory to the pages in which the links are directed.)

**63. Claim 41, 49, and 57 are rejected under 35 U.S.C. 103(a) as being unpatentable over Murashita and Tsuda as applied to claim 39 above, and further in view of Sweet et al. (hereinafter Sweet, US 2002/0073074).**

64. Regarding claim 41, **Murashita and Tsuda** teaches an apparatus substantially as claimed. **Murashita and Tsuda** do not explicitly disclose a maximum link depth designating unit which designates a maximum link depth; and wherein the searching unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

However, **Sweet** teaches a maximum link depth designating unit which designates a maximum link depth; and wherein the searching unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth. (See page 6, paragraph [0063] “One web traversal criterion which may be specified by the user is a maximum depth criterion. This criterion limits the depth of recursive calls to FetchAndIncorporate, and thus limits the ‘link distance’ between the initially retrieved document and subsequently retrieved documents to be incorporated into the target document.”)

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita and Tsuda** with that of **Sweet** because all are related to operating on web documents and by including the maximum link depth as disclosed in **Sweet**, the apparatus can remain efficient by having a limit on the recursion, rather than having unlimited recursion. It is for this reason that one of ordinary skill in the art would have been motivated to include a maximum link depth designating unit which designates a maximum link depth; and wherein the searching unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

65. Regarding claim 49, **Murashita and Tsuda** teaches an apparatus substantially as claimed. **Murashita and Tsuda** do not explicitly disclose a maximum link depth designating unit which designates a maximum link depth; and wherein the searching unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

However, **Sweet** teaches a maximum link depth designating unit which designates a maximum link depth; and wherein the searching unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth. (See page 6, paragraph [0063] "One web traversal criterion which may be specified by the user is a maximum depth criterion. This criterion limits the depth of recursive calls to



FetchAndIncorporate, and thus limits the 'link distance' between the initially retrieved document and subsequently retrieved documents to be incorporated into the target document.”)

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita and Tsuda** with that of **Sweet** because both are related to operating on web documents and by including the maximum link depth as disclosed in **Sweet**, the apparatus can remain efficient by having a limit on the recursion, rather than having unlimited recursion. It is for this reason that one of ordinary skill in the art would have been motivated to include a maximum link depth designating unit which designates a maximum link depth; and wherein the searching unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

66. Regarding claim 57, **Murashita and Tsuda** teach an apparatus substantially as claimed. **Murashita and Tsuda** do not explicitly disclose a maximum link depth designating unit which designates a maximum link depth; and wherein the searching unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

However, **Sweet** teaches a maximum link depth designating unit which designates a maximum link depth; and wherein the searching unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum

Art Unit: 2167

link depth. (See page 6, paragraph [0063] “One web traversal criterion which may be specified by the user is a maximum depth criterion. This criterion limits the depth of recursive calls to FetchAndIncorporate, and thus limits the ‘link distance’ between the initially retrieved document and subsequently retrieved documents to be incorporated into the target document.”)

It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita** and **Tsuda** with that of **Sweet** because the references are related to operating on web documents and by including the maximum link depth as disclosed in **Sweet**, the apparatus can remain efficient by having a limit on the recursion, rather than having unlimited recursion. It is for this reason that one of ordinary skill in the art would have been motivated to include a maximum link depth designating unit which designates a maximum link depth; and wherein the searching unit, in the case where the information could not be extracted from the target document, recursively executes a process for extracting the information from the related document of said document in a range of said designated maximum link depth.

67. **Claims 42 – 44, 50-52, and 58-60 are rejected under 35 U.S.C. 103(a) as being unpatentable over Murashita, Tsuda, and Sweet further in view of Pirolli (hereinafter Pirolli, US 5,895,470).**

68. Regarding claim 42, **Murashita, Tsuda, and Sweet** teach an apparatus substantially as claimed. **Murashita, Tsuda and Sweet** do not explicitly disclose wherein the searching unit discriminates an internal link and an external link on the basis of the document address of the

related document and excludes the documents of the external link from the targets of the information extraction. However, **Pirolli** teaches wherein the searching unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita**, **Tsuda**, **Sweet**, and **Pirolli** because they are related to operating on web documents and by including link discriminating as disclosed in **Pirolli**, the apparatus can be more efficient by only including the pages to search that are likely to be relevant. It is for this reason that one of ordinary skill in the art would have been motivated to include wherein the searching unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

69. Regarding claim 43, **Murashita**, **Tsuda** and **Sweet** teach an apparatus substantially as claimed. **Murashita**, **Tsuda**, and **Sweet** do not explicitly disclose wherein the searching unit executes the information extracting process in order of the document in which a value of the link depth is 3 or fewer. However, **Pirolli** teaches wherein the searching unit executes the information extracting process in order of the document in which a value of the link depth is 3 or fewer. (See

Art Unit: 2167

column 6, line 6, breadth first, e.g. link depth starts at root and continues by level) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita**, **Tsuda**, **Sweet**, and **Pirolli** because they are related to operating on web documents and by including the link depth order as disclosed in **Pirolli**, the apparatus can be more efficient by searching closer links, which usually contain more relevant information, first. It is for this reason that one of ordinary skill in the art would have been motivated to include wherein the searching unit executes the information extracting process in order of the document in which a value of the link depth is 3 or fewer.

70. Regarding claim 44, the combination of **Murashita**, **Tsuda**, **Sweet**, and **Pirolli** additionally discloses wherein the searching unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

71. Regarding claim 50, **Murashita**, **Tsuda** and **Sweet** teach an apparatus substantially as claimed. **Murashita**, **Tsuda**, and **Sweet** do not explicitly disclose wherein the searching unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the

Art Unit: 2167

information extraction. However, **Pirolli** teaches wherein the searching unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita**, **Tsuda**, **Sweet**, and **Pirolli** because they are related to operating on web documents and by including link discriminating as disclosed in **Pirolli**, the apparatus can be more efficient by only including the pages to search that are likely to be relevant. It is for this reason that one of ordinary skill in the art would have been motivated to include wherein the searching unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

72. Regarding claim 51, **Murashita**, **Tsuda**, and **Sweet** teach an apparatus substantially as claimed. **Murashita**, **Tsuda**, and **Sweet** do not explicitly disclose wherein the searching unit executes the information extracting process in order of the document in which a value of the link depth is 3 or fewer. However, **Pirolli** teaches wherein the searching unit executes the information extracting process in order of the document in which a value of the link depth is 3 or fewer. (See column 6, line 6, breadth first, e.g. link depth starts at root and continues by level.) It would have

Art Unit: 2167

been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita, Tsuda, Sweet, and Pirolli** because they are related to operating on web documents and by including the link depth order as disclosed in **Pirolli**, the apparatus can be more efficient by searching closer links, which usually contain more relevant information, first. It is for this reason that one of ordinary skill in the art would have been motivated to include wherein the searching unit executes the information extracting process in order of the document in which a value of the link depth is 3 or fewer.

73. Regarding claim 52, the combination of **Murashita, Tsuda, Sweet, and Pirolli** additionally discloses wherein the searching unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

74. Regarding claim 58, **Murashita, Tsuda, and Sweet** teach an apparatus substantially as claimed. **Murashita, Tsuda, and Sweet** do not explicitly disclose wherein the searching unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. However, **Pirolli** teaches wherein the searching unit discriminates an

internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita**, **Tsuda**, and **Sweet** with **Pirolli** because they are related to operating on web documents and by including link discriminating as disclosed in **Pirolli**, the apparatus can be more efficient by only including the pages to search that are likely to be relevant. It is for this reason that one of ordinary skill in the art would have been motivated to include wherein the searching unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction.

75. Regarding claim 59, **Murashita**, **Tsuda**, and **Sweet** teach an apparatus substantially as claimed. **Murashita**, **Tsuda**, and **Sweet** do not explicitly disclose wherein the searching unit executes the information extracting process in order of the document in which a value of the link depth is small. However, **Pirolli** teaches wherein the searching unit executes the information extracting process in order of the document in which a value of the link depth is small. (See column 6, lines 12 – 26 where the hypertext links are extracting at the higher document depth first, then the links on those pages are executed, finding larger depth value links and then

repeating. In other words, the executing starts with a smaller link depth and then goes to larger link depths during the extraction process.) It would have been obvious to one with ordinary skill in the art at the time of the invention to combine the teachings of **Murashita**, **Tsuda**, and **Sweet** with **Pirolli** because they are related to operating on web documents and by including the link depth order as disclosed in **Pirolli**, the apparatus can be more efficient by searching closer links, which usually contain more relevant information, first. It is for this reason that one of ordinary skill in the art would have been motivated to include wherein the searching unit executes the information extracting process in order of the document in which a value of the link depth is small.

76. Regarding claim 60, the combination of **Murashita**, **Tsuda**, **Sweet**, and **Pirolli** additionally discloses wherein the searching unit discriminates an internal link and an external link on the basis of the document address of the related document and excludes the documents of the external link from the targets of the information extraction. (See **Pirolli**, column 6, lines 12 – 15 “The returned page is then parsed to extract hyperlinks to other pages, step 202. Links that point to pages within the Web locality are added to a list of pages to request and retrieve.” Here, the pages that are not in the web locality are not added to the list, thereby discriminating internal and external links.)

#### ***Response to Arguments***

77. Applicant's arguments with respect to claims 1-8, 10-46, and 48-65 have been considered but are moot in view of the new ground(s) of rejection. Applicant's assert the following (lettered):



Art Unit: 2167

A. That Pirolli and tsuda do not disclose or suggest from amended claim 1:

a category layer specifying unit in which the category of the information to be extracted is expressed by a layer structure;

an extracting unit which:

in case where only an extraction result of a lower layer in said layer structure exists and an extraction result of an upper layer is missing as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, extracts a character string of a layer which is higher than that of the extraction result lower layer from the related document of said target document.

*In response, the examiner respectfully disagrees.*

*A category layer specifying unit in which the category of information to be extracted is expressed by a layer structure is disclosed by Tsuda. As Tsuda, col. 7 lines 15-17, discloses a directory file creating unit (e.g. category layer specifying unit) creates hyper text format directory (i.e. layer structure) file with the data 56 (i.e. category of information), the data 33 (i.e. category of information), the data 52 (i.e. category of information), the data 53 (i.e. category of information), the data 54 (i.e. category of information), and the data 55 (i.e. category of information). Accordingly, Applicant's arguments directed towards a category layer specifying*

Art Unit: 2167

*unit in which the category of the information to be extracted is expressed by a layer structure, are unpersuasive.*

*An extracting unit which, in case where only an extraction result of a lower layer in said layer structure exists and an extraction result of an upper layer is missing as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, extracts a character string of a layer which is higher than that of the extraction result lower layer from the related document of said target document is disclosed by Tsuda. As Tsuda discloses the col. 15 lines 21-23 that the directory file creating unit designates (i.e. designated by said start point address designating unit) a set of keywords registered in the field down (i.e. corresponding to the category from the target document) of the keyword w (target document) of the keyword table 62 as S2. Col. 15 lines 24-29, the directory file creating unit 43 determines whether or not s2 is empty (i.e. extraction result of a lower layer in said layer structure exists and an extraction result of an upper layer is missing as a result of the extraction of the information). When s2 is not empty the directory file creating unit extracts a keyword from s2 (i.e. extracts a character string layer which is higher than that of the extraction result lower layer from the related document of said target document). Next the directory file creating unit determines whether or not the path field of the keyword u is empty.*

*Accordingly, as shown above Tsuda suggests the above claimed elements; therefore, applicant's assertions directed towards*

*“a category layer specifying unit in which the category of the information to be extracted is expressed by a layer structure;*

*an extracting unit which:*

*in case where only an extraction result of a lower layer in said layer structure exists and an extraction result of an upper layer is missing as a result of the extraction of the information corresponding to the category from the target document designated by said start point address designating unit, extracts a character string of a layer which is higher than that of the extraction result lower layer from the related document of said target document.” In claim 1, are unpersuasive over the cited references.*

B. That Tsuda does not make use of the full term “Dr. Akiyama’s laboratory” without being told to use the full term as keyword (col. 5 l. 47-52, tsuda). Exemplary examples are provided in specification page 28 and 29.

*In response, the examiner respectfully states that examples are not definitions nor limit the claims to any particular scope. Instead, merely provide one use of the claimed inventive steps and features.*

C. That Tsuda, keywords are used to hierarchically organize and establish relationships between the documents. That keywords are not used to extract information from the documents.

Additionally, a keyword as used in Tsuda is not a category expressed by a layer structure. That

Art Unit: 2167

Keywords in Tsuda may be arranged hierarchically, but they may not define categories to be searched for in the hypertext document.

*In response, the examiner respectfully disagrees with applicant's assertions. Secondly, none of the arguments presented above do not appear to directly correspond to the claimed limitations. These arguments simply appear to be explaining the differences in broad terms of the application and the cited reference.*

*However, with respect, to keywords are not used to extract information from the documents. The examiner respectfully disagrees because as seen in figure 2, the keywords are extracted from documents. That element 41 of figure 2 trims the words out of the document 21 and are further used in the processing unit 11. Hence, keywords are used to extract information from the documents.*

*As to a keyword as used in Tsuda is not a category expressed by a layer structure. The examiner respectfully disagrees because Tsuda also discloses, col. 3 lines 22-24, associative relations as a link is added to directory information (i.e. layer structure) that that represents categorized results of a group of documents.*

*As to keywords in Tsuda may be arranged hierarchically, but they may not define categories to be searched for in the hypertext. The examiner respectfully disagrees because Tsuda, col. 1 lines 62-65 discloses hierarchical categorizes are created with keywords of a group of documents and each document is registered in a plurality of categories.*

D. That claims 47 and 55 have been incorporated into claim 39, and are allowable over similar reasons as claim 1 above. Essentially, that Tsuda fails to teach the features of amended claim 39 which are similar to amended claim 1.

*In response, please see above A and C.*

*Conclusion*

78. The prior art made of record listed on PTO-892 and not relied, if any, upon is considered pertinent to applicant's disclosure.

79. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire **THREE MONTHS** from the mailing date of this action. In the event a first reply is filed within **TWO MONTHS** of the mailing date of this final action and the advisory action is not mailed until after the end of the **THREE-MONTH** shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than **SIX MONTHS** from the date of this final action.

*Contact Information*

80. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Michael D. Pham whose telephone number is (571)272-3924. The examiner can normally be reached on Monday - Friday 9am - 5:00pm.

Art Unit: 2167

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, John Cottingham can be reached on 571-272-7079. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Michael Pham  
Art Unit 2167  
Examiner

*M.F.*

Cam Y. Truong  
Art Unit 2162  
Primary Examiner

*cy*

John Cottingham  
Art Unit 2167  
Supervisor

  
JOHN COTTINGHAM  
SUPERVISORY PATENT EXAMINER  
TECHNOLOGY CENTER 2100